(19) FEDERAL REPUBLIC OF GERMANY

GERMAN PATENT OFFICE

(12) **Patent application laid open**

(51) Int. Cl.[5]:
C 12 Q 1/68

(10) **DE 41 41 178 A1**

G 01 N 33/68
C 07 H 21/04
//C12N 9/22

(21) File number: P 41 41 178.1
(22) Application date: 13 Dec. 91
(43) Date laid open: 17 June 93

(71) Applicant:

Europäisches Laboratorium für Molekularbiologie (EMBL), 6900 Heidelberg, DE

(74) Attorney:

Weickmann, H., Dipl.-Ing.; Fincke, K., Dipl.-Phys. Dr.; Weickmann, F., Dipl.-Ing.; Huber, B., Dipl.-Chem.; Liska, H., Dipl.-Ing. Dr.-Ing.; Prechtel, J., Dipl.-Phys. Dr.rer.nat.; Böhm, B., Dipl.-Chem. Univ. Dr.rer.nat., Patent Attorneys, 8000 Munich

(72) Inventor:

Ansorge, Wilheim, Dr., 6901 Gaiberg, DE

(54) Method for sequencing of nucleic acids

(57) For the sequencing of nucleic acids, one sequences a complimentary strand, working step by step, to an at least partially single-stranded nucleic acid being sequenced as a template strand, by means of a suitable polymerase and all four nucleotides, starting from a primer or a double-stranded segment, using labeled nucleotides and performing the following steps:
(1) Incubation of the nucleic acid provided with the primer or being partially double-stranded with an incubation mixture consisting of a polymerase, one of the four types of nucleotides in labeled form, and other substances required for the chain polymerization, wherein a labeled nucleotide is incorporated into the growing complementary strand if the next nucleotide available on the template strand is complementary to the labeled nucleotide used.
(2) Separation of the nucleic acid, possibly longer by one nucleotide, from the incubation mixture of Step (1) and performance of one or more washing steps in familiar fashion.
(3) Determination of the incorporation of a nucleotide by means of its labeling, and
(4) if the incorporation of a labeled nucleotide has occurred, elimination of the labeling,
wherein Steps (1) through (4) are carried out in cycles for each of the four types of nucleotides.

Specification

The invention concerns a method for sequencing of nucleic acids.

Sequencing of nucleic acids today numbers among the daily routine chores in biochemistry laboratories. Normally, sequencings are performed by one of the two standard sequencing techniques, namely, either the chemical decomposition method of Maxam-Gilbert or the enzymatic complementary strand synthesis method of Sanger. Presently known automated sequencing methods per the Sanger method use, for the labeling, either end-labeled primers or labeled terminator-ddNTPs. In the standard sequencing techniques, the labeled fragments are separated in terms of size on a medium, such as a polyacrylamide gel, and determined via the labeling. The labeling in this case is radioactive labeling, fluorescent labeling, etc. The separating capacity of the gel matrix is a limiting factor on the resolution and the length of the DNA sequence which can be determined.

Distinct limits are set by the systems known thus far. But since sequencing is performed so often, improvements over the known methods are certainly desirable.

Thus, the purpose of the present invention is to provide a method for the sequencing of nucleic acids in which even very long nucleic acid sequences can be determined and in which even very small amounts of nucleic acid lead to definite results.

This purpose is accomplished by the invented method for sequencing of nucleic acids in which one synthesizes a complementary strand to an at least partially single-stranded nucleic acid being sequenced as a template strand, step by step, using a suitable polymerase and all four nucleotides, starting with a primer or a double-stranded segment, wherein one uses labeled nucleotides, and which comprises the following steps:

(1) Incubation of the nucleic acid provided with the primer or being partially double-stranded with an incubation mixture consisting of a polymerase, one of the four types of nucleotides in labeled form, and other substances required for the chain polymerization, wherein a labeled nucleotide is incorporated into the growing complementary strand if the next nucleotide available on the template strand is complementary to the labeled nucleotide used.
(2) Separation of the nucleic acid, possibly longer by one nucleotide, from the incubation mixture of Step (1) and performance of one or more washing steps in familiar fashion.
(3) Determination of the incorporation of a nucleotide by means of its labeling, and
(4) if the incorporation of a labeled nucleotide has occurred, elimination of the labeling,

wherein Steps (1) through (4) are carried out in cycles for each of the four types of nucleotides.

The entirely new approach which has led to the invented method is that only one kind of nucleotide, and this in labeled form, is presented for the formation of the complementary strand. Thus, if the next base occurring on the template strand is complementary to the labeled nucleotide present in the incubation mixture, the nucleotide along with its labeling is incorporated by the polymerase. In the context of the invention, one shall use a polymerase that is capable of incorporating the labeled nucleotide into the

growing complementary strand. Depending on the kind of nucleic acid being determined, one can thus use here the familiar polymerases, such as DNA-polymerase – T7 DNA-polymerase, for example. The nucleic acids being determined are not only DNA fragments, but also RNAs. In the event that a labeled nucleotide is incorporated, this will be established by the labeling, thus allowing the conclusion as to a complementary base in the template strand.

If the base on the template strand is not complementary to the labeled nucleotide which is presented, no incorporation will occur and no labeling signal will be obtained.

The invented method is furthermore characterized by the subsequent removal of the labeling in the event that a labeled nucleotide has been incorporated. By "removal of the labeling" in the context of the invention is meant that the nucleic acid is treated in such a way that the label signal disappears. However, this does not mean that a chemical splitting off of the labeling part of the molecule must occur; a bleaching with a strong laser is also conceivable in the case of a fluorescent labeling (for example, see B. Scalettar et al., *Biophys. J.* 53, 215 (1988) or Mathies, R.A., Stryer, L., "Single-Molecule Fluorescence Detection: A Feasibility Study using Phycoerythrin." *Applications of Fluorescence in the Biomedical Sciences*, p. 129-140 (1986) Alan R. Liss. Inc.). In any case, however, it is essential that the removal of the labeling during Step (4) does not chemically modify the labeled nucleotide in a way that prevents the polymerizing of the next nucleotide complementary to the template strand.

For example, when determining a DNA sequence in the invented method, if the next base incorporated after the primer or partial double strand is a labeled dATP, the label signal which differs from the control signal of the DNA without labeling reveals that the base T is present for the template strand in this position. If several identical bases are present on the template strand, i.e., thymidine in this case, a corresponding number of adenosine molecules will be incorporated in the complementary strand, yielding a correspondingly stronger label signal.

The same step that has been described as an example for the labeled nucleotide adenosine in the incubation mixture of Step (1) is then carried out in cycles for all kinds of nucleotides. Thus, for each base on the template strand, the complementary labeled nucleotide will be presented and incorporated into the chain, and because of the removal of the labeling after each particular incorporation only the last incorporated labeled nucleotide will produce a signal. For this reason, the label signal is not cumulative and the accuracy of the method is very great.

Figure 1 shows the sequence of the method by means of a flow chart.

In the context of the invention, it is preferable to use a single-stranded nucleic acid for the sequencing and to hybridize an oligonucleotide as primer in the 5'-3'-direction in front of the sequence being made.

In a preferred embodiment of the invention, the nucleic acid being sequenced is bound to a solid phase before Step (1). This makes it possible to easily dip the nucleic acid into the solutions and mixtures required for the various steps and also to easily separate it once again from these solutions, especially in automated systems.

It is important that the solid phase produce little or no background for the label signal, e.g., if fluorescence is used as the label, the solid phase must consist of a material that does not fluoresce in the same emission wavelength range as the fluorescence label of the particular nucleotides.

The anchoring of the nucleic acid to the solid phase can be done by familiar methods; preferably the anchoring is done via a specific binding pair, one partner of which is joined to the solid phase, and the other partner is bound to the nucleic acid. Especially preferred here as the specific binding pair in the context of the invention is the system biotin/avidin or biotin/streptavidin, wherein it is again preferable that the avidin or streptavidin be bound to the solid phase in familiar fashion and the nucleic acid is modified by biotin.

Suitable as the solid phase are all materials which are inert relative to the substances in the solutions used, which enable a fixation of the nucleic acid, and which (as mentioned above) do not differ with the labeling. Preferred examples which can be mentioned here are glass, a polymer membrane, or polymer or glass beads.

In another preferred embodiment of the invention, instead of nucleic acids in a fixed form bound to a solid phase one uses a flow system for Steps (1) through (4), wherein the nucleic acid remains in solution and is separated from the other particular substances used by means of filters and/or capillaries.

Preferred labeling in the nucleotides according to the invention is a fluorescent labeling. In particular, one can use fluorescent labels as are described in German patent application P 41 25 745. Also, the so-called "time resolved fluorescence" can be advantageously used as labeling.

In the context of the invention, the nucleotides used are deoxyribonucleotides. In another preferred embodiment of the invention, however, one can also use dideoxyribonucleotides or so-called "caged" nucleotides (see, for example, *Handbook of Fluorescent Probes and Research Chemicals*. Richard P. Haugland, 1989, Molecular Probes Inc., Eugene, USA or Matthews and Kricka. *Analyt. Biochem.* 169, 1 (1988)). A nucleotide in which the fluorescent dye is connected to the 3'OH-group also appears suitable (Krayevsky, A., BBA, 1986, 868, p. 136). It turns out that each time only one dideoxyribonucleotide can be incorporated, since the polymerase no longer has a capability of extending the chain further, due to these so-called terminators, even when the nucleotide corresponding to the next base on the template strand is present. This means that even when identical bases occur several times in succession on the template strand, only one nucleotide is incorporated during each incubation step. This can contribute to the accuracy of the method. After determining the labeling, the dideoxyribonucleotide is then converted into a "dNTP-like polymerization-extendible nucleotide". In this way, the termination of the synthesis of the complementary strand is canceled and the next cycle of Steps (1) through (4) can ensue. However, thanks to the use of dideoxyribonucleotides, an even more simplified method is possible, wherein all four kinds of nucleotides are used as dideoxyribonucleotides, yet the four different dideoxys each have different fluorescent labels. Which of the dideoxyribonucleotides has in fact been incorporated into the growing complementary strand can then be established in Step (3) of the invented method by means of the particular label. This preferred embodiment of the invented method therefore signifies another substantial facilitation and acceleration of the sequencing method.

In the context of the invention, it is preferable to denature the nucleic acid prior to the sequencing. Furthermore, it is preferable, each time after Step (3) or (4), to fill in any gaps in the complementary strand with a nonlabeled nucleotide corresponding to the labeled nucleotide used in Step (1). This measure should also further improve the

accuracy of the method. Since, of course, not just one DNA molecule is determined in the invented method, but rather many of them are required for the determination and the producing of a detectable signal, it may also happen that the correct incorporation of a nucleotide is mistakenly omitted in many strands. And in turn, the next nucleotide also cannot be incorporated thereafter, so that these nucleic acids would be lost to the subsequent signalling. Therefore, any defective locations which occur are again filled in with a nonlabeled nucleotide, so that the same starting conditions exist for all nucleic acids for the next cycle of Steps (1) through (4) with the next nucleotide.

When using a solid phase-bound nucleic acid for the sequencing, it is also possible to determine several, indeed, very many nucleic acids at the same time. For this, the nucleic acids being sequenced are anchored to a solid phase in spatial separation and they are treated at the same time with the various incubation mixtures and solutions. One can produce a kind of microstructure on reduced scale and sequence hundreds or thousands of DNA samples on such a "DNA chip" on microscopic scale at the same time (see Figure 2). The DNA can be anchored to the solid phase by a kind of automated and computer-controlled microcellular injection. But all other known ways of anchoring are also possible here. Furthermore, the nucleic acid can also be anchored to the bottom of a microtitration plate and the solutions and incubation mixtures can be placed in the cavities and removed once again.

As already mentioned above, it is especially preferable in the context of the invention to carry out the sequencing in an automatic machine. In particular, a Charge Coupled Device (CCD) camera can be used to determine the labeling.

The invented method, which represents a totally new technique for the sequencing of DNA, without the need for separation on agarose or polyacrylamide gels, represents a fast and extremely accurate possibility of sequencing nucleic acids. It is even possible to determine very small quantities of nucleic acids, many times smaller than present-day methods, since each DNA molecule contributes to the label signal, whereas in the methods thus far a statistical distribution of fragments of nucleic acid and, thus, only a fraction of the starting quantity of nucleic acids is available for each nucleotide being determined.

The accuracy of the method is based on the fact that the extension of the polymerized chain is interrupted as soon as the polymerase enzyme arrives at a base on the template strand that is not present in the incubation mixture with the labeled nucleotides that is being used to incubate the nucleic acid. Although in theory it is possible for a base to be incorporated wrong, which would produce a slight background signal, the wrong labeling is, nevertheless, also eliminated after the label removal Step (4), so that the wrong signal, which might impair the resolution after several cycles, cannot accumulate. For this reason, the background is very low in the invented method and the accuracy of the sequence determination is correspondingly high.

In order to enhance the accuracy of the sequence determination even more, the nucleic acid sample being determined can be separated into four subsamples and then all steps of the invented method are carried out with each of the four portions of nucleic acid. The sequence will then be determined not only in terms of the positive signal from the particular sample in which a labeled nucleotide has been incorporated, but also in terms of the absence of a signal from the samples which have been incubated with the other three labeled nucleotides.

In summary, the invented method offers a fast and easily automated way of reliably establishing with great accuracy the sequence of nucleic acids, even when only very small quantities of such are available.

The following figures further explain the invention:

Figure 1 shows a flow chart of the invented method.

Figure 2 shows the arrangement of many different nucleic acids, here: DNA molecules, on a solid phase on microscopic scale for simultaneous processing.

Figure 3 shows the detection of the incorporation of fluorescein-12-dUTP by Klenow-DNA-polymerase.

Patent claims

1. Method for sequencing of nucleic acids, in which one synthesizes a complementary strand to an at least partially single-stranded nucleic acid being sequenced as a template strand, step by step, using a suitable polymerase and all four nucleotides, starting with a primer or a double-stranded segment, wherein one uses labeled nucleotides, and which comprises the following steps:

(1) Incubation of the nucleic acid provided with the primer or being partially double-stranded with an incubation mixture consisting of a polymerase, one of the four types of nucleotides in labeled form, and other substances required for the chain polymerization, wherein a labeled nucleotide is incorporated into the growing complementary strand if the next nucleotide available on the template strand is complementary to the labeled nucleotide used.

(2) Separation of the nucleic acid, possibly longer by one nucleotide, from the incubation mixture of Step (1) and performance of one or more washing steps in familiar fashion.

(3) Determination of the incorporation of a nucleotide by means of its labeling, and

(4) if the incorporation of a labeled nucleotide has occurred, elimination of the labeling,

wherein Steps (1) through (4) are carried out in cycles for each of the four types of nucleotides.

2. Method per Claim 1, characterized in that the nucleic acid being sequenced is anchored to a solid phase before Step (1).

3. Method per Claim 2, characterized in that the anchoring is done by a specific binding pair, one partner of which is joined to the solid phase, and the other partner is bound to the nucleic acid.

4. Method per Claim 3, characterized in that one uses, as the specific binding pair, the system biotin/avidin or biotin/streptavidin.

5. Method according to one of Claims 2 to 4, characterized in that the solid phase used is glass, a polymer membrane, or polymer or glass beads.

6. Method according to one of the preceding claims, characterized in that one hybridizes an oligonucleotide as the primer in front of the sequence being sequenced.

7. Method per Claim 1 or 6, characterized in that one uses, for Steps (1) through (4), a flow system in which the nucleic acid remains in solution and it is separated from other substances by means of filters and/or capillaries.

8. Method according to one of the preceding claims, characterized in that one uses a fluorescent labeling as the labeling of the nucleotides.
9. Method according to one of the preceding claims, characterized in that one uses deoxyribonucleotides as the nucleotides.
10. Method according to one of Claims 1 through 8, characterized in that one uses dideoxyribonucleotides as the nucleotides and converts them into deoxyribonucleotide-like polymerization-extendible nucleotides after the determination of the labeling is done in Step (3) or possibly after Step (4).
11. Method per Claim 11, characterized in that one uses all four kinds of nucleotides, each with different fluorescent labels, together in an incubation mixture in Step (1), and the distinguishing of the incorporated nucleotides is done in terms of their labeling.
12. Method according to one of the preceding claims, characterized in that one denatures the nucleic acid prior to the sequencing.
13. Method according to one of the preceding claims, characterized in that one fills up any gaps existing in the complementary strand after Step (3) or (4) with a nonlabeled nucleotide corresponding to the labeled nucleotide used in Step (1).
14. Method according to one of Claims 1 through 6 and 8 through 13, characterized in that one anchors several nucleic acids being sequenced in a spatial separation from each other on the solid phase and sequences them simultaneously.
15. Method according to one of the preceding claims, characterized in that one carries out Steps (1) through (3) and possibly also (4) in an automatic system.
16. Method according to one of the preceding claims, characterized in that one uses a Charge Coupled Device (CCD) camera to determine the labeling.

Plus 3 pages of drawings

[Keys to the figures:]

Figure 1

a. SCHEMATIC REPRESENTATION OF THE DETERMINATION OF DNA AND RNA SEQUENCES ACCORDING TO THE INVENTION

b. SOLID PHASE

c. SOLID PHASE-BOUND DNA (SINGLE-STRANDED)

d. LABELED dNTP=dNTP*

e. STEP 2

f. VESSEL WITH dATP* + ENZYME + BUFFER

g. WASHING

h. DETECTION OF A SIGNAL BY THE LABEL

i. ANALYZER

j. BLEACHING

k. YES SIGNAL

l. NO SIGNAL

m. AS IN THE PRECEDING STEPS 2-5

Figure 2

a. SIMULTANEOUS TREATMENT (SEQUENCING) OF VARIOUS (MANY HUNDREDS OF) DNA SAMPLES, ESPECIALLY ON MICROSCOPIC SCALE

b. SOLID SUBSTRATE (E.G., MICROTITRATION PLATE OR MEMBRANE)

c. MANY DIFFERENT DNA SAMPLES, BOUND TO THE SURFACE OF THE SOLID SUBSTRATE, ARE ALL TREATED SIMULTANEOUSLY IN UNIFORM, PERIODIC MANNER AND THEIR SEQUENCES ARE THUS OBTAINED AT THE SAME TIME.

d. ON MICROSCOPIC SCALE, DIMENSIONS OF SEVERAL MILLIMETERS

Fig.1

a. SCHEMATISCHE DARSTELLUNG DER ERFINDUNGSGEMÄSSEN BESTIMMUNG VON DNA- UND RNA-SEQUENZEN

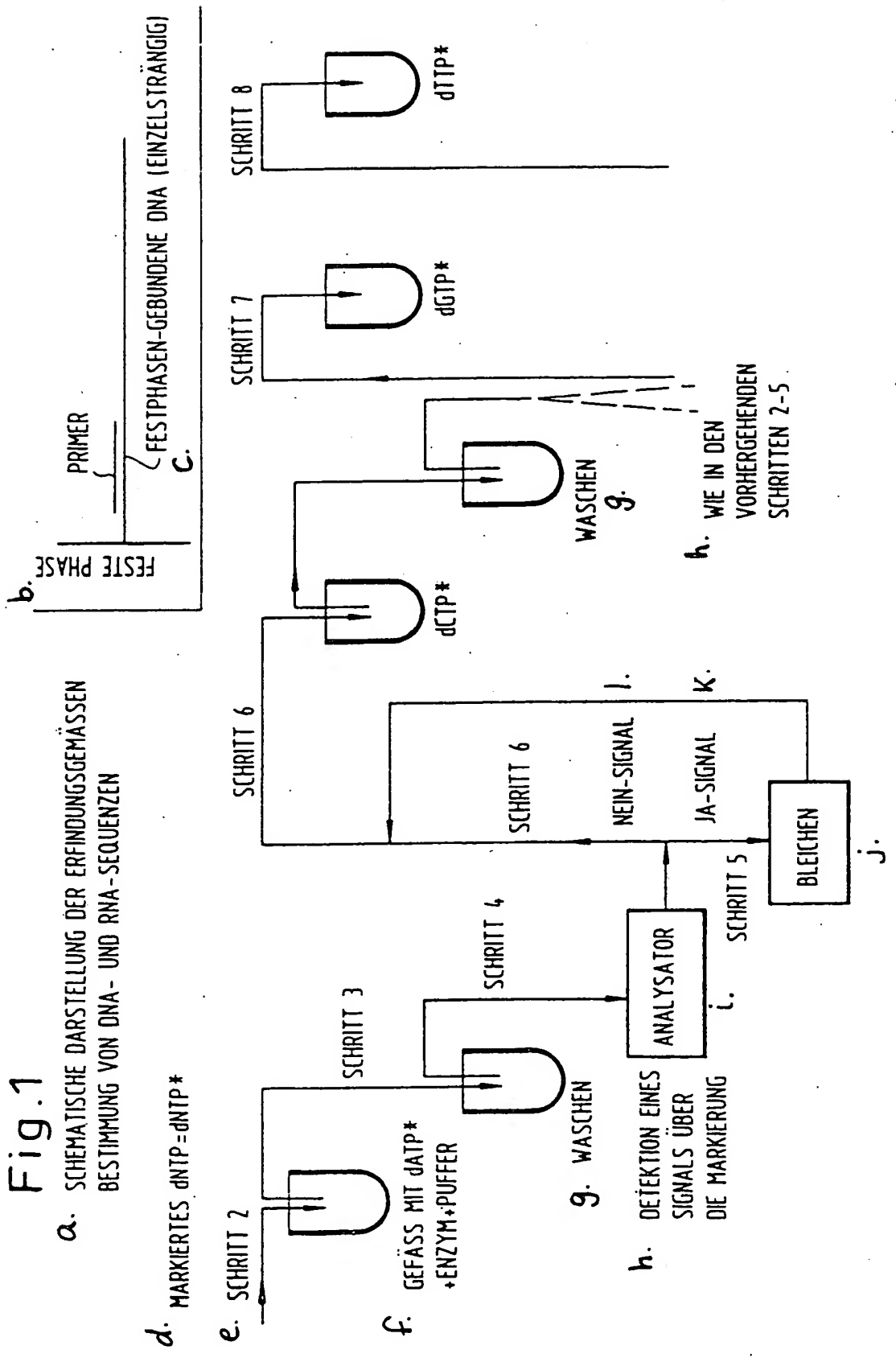

Fig. 2

a. GLEICHZEITIGE BEHANDLUNG (SEQUENZIERUNG) VERSCHIEDENER (VIELER HUNDERTER) DNA-PROBEN, INSBESONDERE IM MIKROMASSTAB.

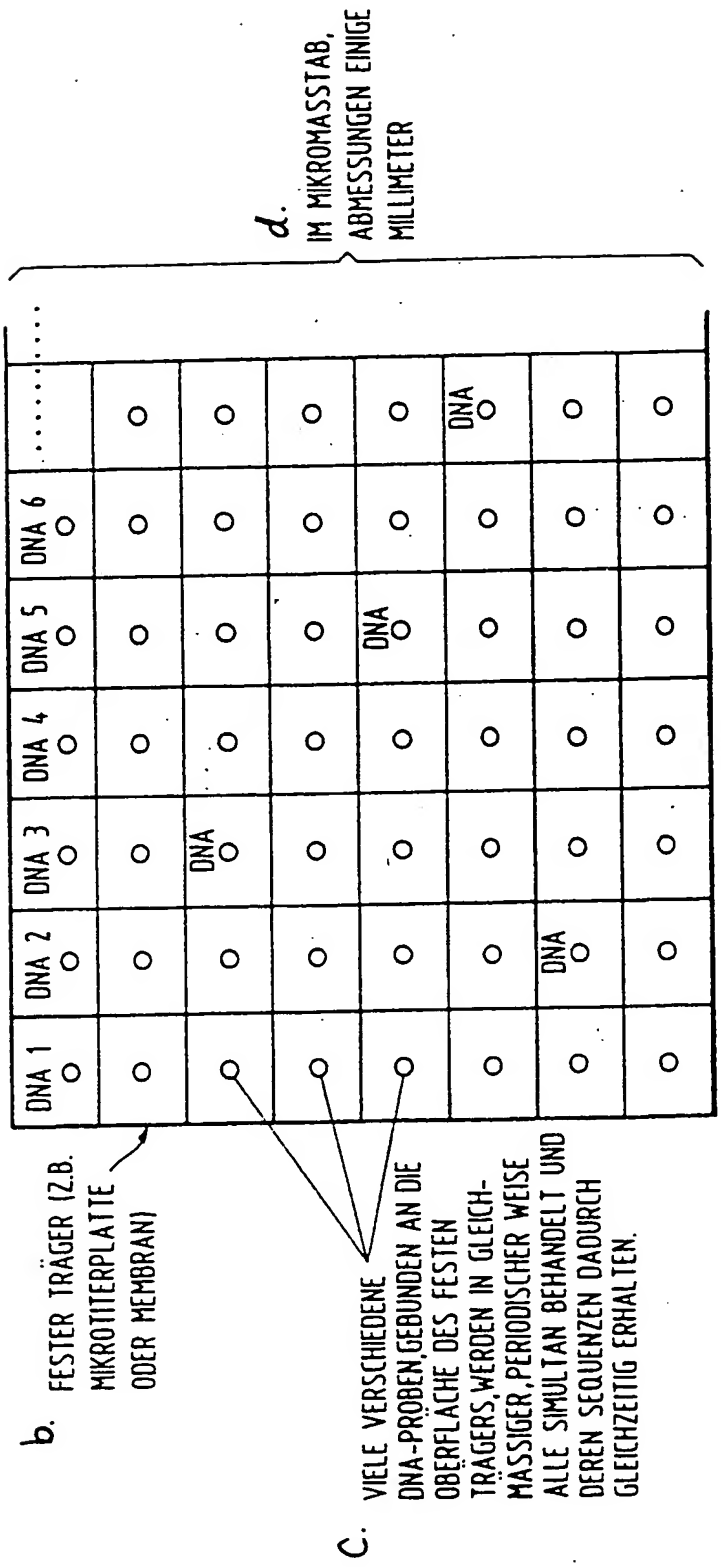

Figure 3

ENZYMATIC INCORPORATION OF FLUORESCEIN-12-dUTP BY KLENOW-DNA-POLYMERASE. THE DETECTION INVOLVES AROUND 10^{-17} MOLE OF THE LABELED DNA FRAGMENT